

EMILiA

Entwicklung einer E-Mail-Archivierungssoftware für
kulturelle Gedächtnisinstitutionen

Elisabeth Klindworth (Archiv der Max-Planck-Gesellschaft)

Nico Beyer (Freie Universität Berlin)



Freie Universität  Berlin



ARCHIV
DER MAX-PLANCK-GESellschaft

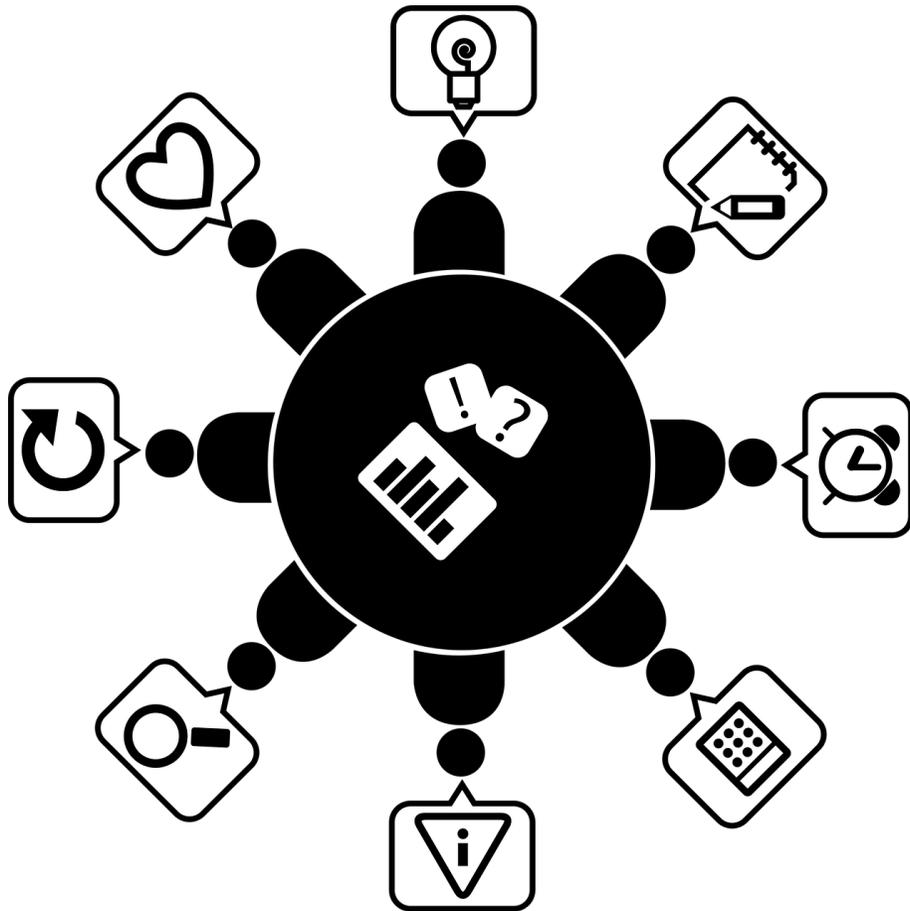
Themen

- 1 Rückblick
- 2 Ausgangspunkt
- 3 Aktueller Stand
- 4 Herausforderungen
- 5 Zeit für Fragen

1

Rückblick

Rückblick



- 2015: Beginn Konzeption in Kooperation mit Fachbereich Informatik der FU Berlin
- Einstellung studentische Hilfskraft im Archiv der Max-Planck-Gesellschaft
- Entwicklung Spezifikation und Prototyp
- Tests mit Echtdateien des Archivs der Max-Planck-Gesellschaft
- 2024: Projektförderung durch „ProValid“ der Investitionsbank Berlin

Ausgangspunkt

2

Ausgangspunkt



Spam, Werbung, E-Mails ohne Informationswert



Händische Bewertung und Erschließung ist aufgrund von riesigen Datenmengen nicht zu bewerkstelligen



Personenbezogene Daten



Viele E-Mails müssen anonymisiert werden



Schadhafte Anhänge



E-Mails können Viren enthalten

3

Aktueller Stand

Grundkonzept

Erfassung, **M**anagement, **I**ndizierung, **L**imitierung, **i**ntelligente **A**nalyse



Abgabe an das
Archiv



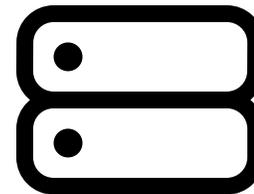
Übernahme

- Import
- Virenprüfung
- Formaterkennung
- Authentizität
- Integrität
- Erstellung von SIPs



Bewertung &
Erschließung

- Erkennung von Spam
und Dubletten
- Erkennung von
Themen, Personen
und Orten
- Identifikation
personenbezogener
Daten



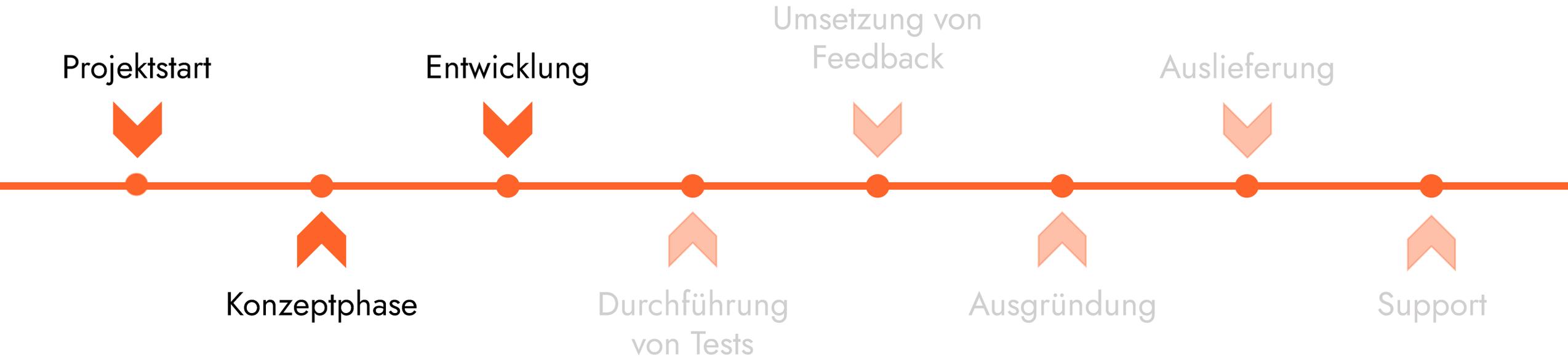
Übergabe an
dLZA



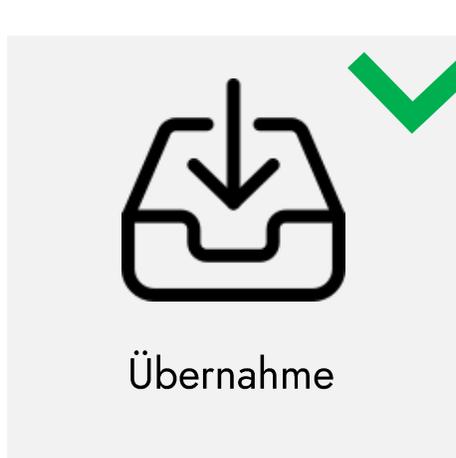
Nutzung

- Recherchedatenbank
- Anonymisierung
- Erstellung von DIPs
- Darstellung

Meilensteine



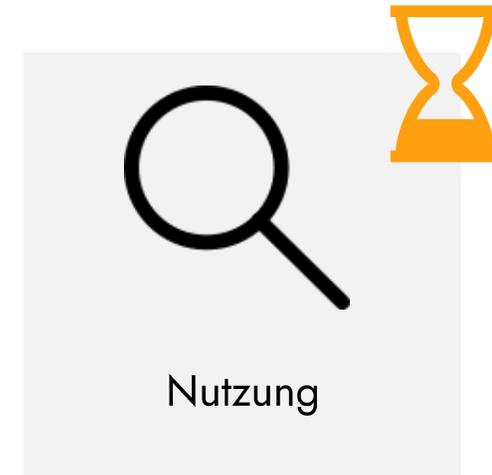
Entwicklungsstand



- Übernahme und Erstellung von SIPs funktioniert bereits



- Erkennung von Klarnamen, Sprachen, Themen und Threads wurde getestet aber noch nicht implementiert
- Dubletten- und Spamererkennung müssen noch evaluiert werden



- Recherchedatenbank und Viewer müssen noch entwickelt werden

5

Herausforderungen

Herausforderungen

1

E-Mail Standard ist sehr offen, weshalb die Metadaten sehr unterschiedlich aussehen können

2

Unterschiedliche E-Mail-Clients, proprietäre Dateiformate und Alter der Daten können zu Defekten und Anomalien führen

3

Umgang mit verschlüsselten E-Mails ist noch unklar

4

Unterschiedliche Zeichenkodierungen können dazu führen, dass einige E-Mails nicht korrekt dargestellt werden

5

Entscheidung für ein archivtaugliches Containerformat steht noch aus

6

Test an Echtdaten zeigt aktuell, dass die Informationsdichte sehr gering ist

6

Zeit für Fragen

Vielen Dank für Ihre Aufmerksamkeit!

Fragen und Vorschläge sind jederzeit willkommen.

Nico Beyer

EMILiA-Projekt

E-Mail: nico.beyer@emilia-archiv.de

Telefon: +49 30 / 841 337 15

Boltzmannstraße 14, 14195 Berlin-Dahlem

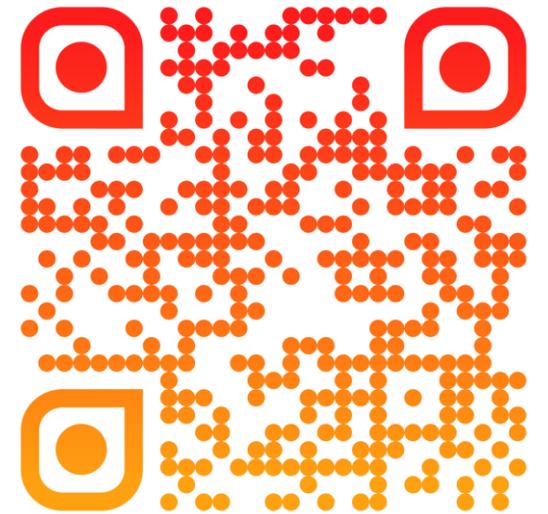
Elisabeth Klindworth

Archiv der Max-Planck-Gesellschaft

E-Mail: klindworth@archiv-berlin.mpg.de

Telefon: +49 30 / 841 337 12

Boltzmannstraße 14, 14195 Berlin-Dahlem



www.emilia-archiv.de

